THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 276, No. 30, Issue of July 27, pp. 28372–28379, 2001
Printed in U.S.A.

NMR Solution Structure of Murine CCL20/MIP-3α, a Chemokine That Specifically Chemoattracts Immature Dendritic Cells and Lymphocytes through Its Highly Specific Interaction with the β-Chemokine Receptor CCR6*

Received for publication, April 9, 2001, and in revised form, May 22, 2001
Published, JBC Papers in Press, May 23, 2001, DOI 10.1074/jbc.M103121200

José Manuel Pérez-Cañadillas‡, Ángel Zaballos§, Julio Gutiérrez§, Rosa Varona§, Fernando Roncal§, Juan Pablo Albar§, Gabriel Márquez§, and Marta Bruix‡¶

From the ‡Instituto de Estructura de la Materia, Consejo Superior de Investigaciones Científicas, Serrano 119, 28006 Madrid, Spain and the §Departamento de Inmunología y Oncología, Centro Nacional de Biotecnología, Campus Universidad Autónoma de Madrid, Cantoblanco, 28040 Madrid, Spain

CCL20/MIP-3α is a β-chemokine expressed in the thymus, skin, and intestinal epithelial cells that exclusively binds and activates the CCR6 receptor in both mice and humans. The strict receptor binding specificity of CCL20 is exceptional; other chemokines and their receptors bind promiscuously with multiple partners. Toward determining the structural basis for the selective receptor specificity of CCL20, we have determined its three-dimensional structure by ^{1}H NMR spectroscopy. CCL20 exhibits the same monomeric structure previously described for other chemokines: a three-stranded β-sheet and an overlying α-helix. The CCL20 receptor selectivity could arise from the rigid conformation of the N-terminal DCCL motif as well as the groove between the N-loop and the β_2-β_3 hairpin, which is significantly narrower in CCL20 than in other chemokines. Similar structural features are seen in human β-defensin 2, a small nonchemokine polypeptide reported to selectively bind and activate CCR6, which stresses their importance for the specific binding of both CCL20 and β-defensin 2 to CCR6. CCL20's structure will be useful to design tools aimed to modulate its important biological functions.

Chemokines are a large family of nearly 50 small chemotactic cytokines that regulate the trafficking of leukocytes and the host inflammatory responses by specifically interacting with G protein-coupled receptors (1–4). Based on conserved cysteine residues in the N terminus of the molecules, CXC, CC, C, and CX3C chemokine subfamilies have been described (1, 4). The hallmark of chemokine function is considered to be the regulation of leukocyte trafficking, but chemokines also play important roles in the control of angiogenesis, tumor metastasis, hematopoiesis, and human immunodeficiency virus infection (1–5). These properties make chemokines and their receptors very interesting targets for therapeutic intervention.

Eighteen different chemokine receptors have been described so far (6). Their interactions with their ligands are clearly characterized by promiscuity, since most chemokine receptors are able to bind several chemokines and many chemokines can interact with several receptors. CCL20/MIP-3α, a CC chemokine, is an exception to this general rule. In both humans and mice, CCL20 binds only to CCR6 receptor (7, 8), which is not bound or activated by any other chemokine. Besides its strict receptor binding specificity, CCL20 has an interesting expression pattern. Murine CCL20 is expressed in the thymus (8), epithelial cells from intestinal tissue (9), and the follicle-associated epithelium in Peyer's patches (10). In addition, data reported on human and mouse CCR6 show the specific expression of this receptor in different dendritic cell and lymphocyte subsets (8, 10–13).

Taken together, these data suggest a role for CCR6/CCL20 signaling in regulating the migration and recruitment of antigen-presenting and immunocompetent cells during inflammatory and immunological responses. Indeed, two recent studies on $CCR6^{-/-}$ mice showed that these animals have defects in leukocyte homing to the intestinal mucosa, impaired humoral immunity, and altered responses in contact hypersensitivity and delayed type hypersensitivity models of inflammation (14, 15). Strikingly, a recent study showed that human β-defensins, small polypeptides that contribute to host defense and whose overall amino sequence shows no homology to CCL20, are able to recruit dendritic and T cells by interacting only with CCR6, out of several different chemokine receptors tested (16).

Detailed structural studies on different chemokines have been performed to help to identify the motifs that are key for the specificity of receptor binding and activation (17–36). These studies have been carried out on promiscuous chemokines with different degrees of receptor binding specificity, whose receptors are also able to bind some other chemokine(s). As an approach to study the basis of the strict binding specificity of CCL20 to CCR6, we have investigated the conformation of CCL20 by two-dimensional ^{1}H NMR spectroscopy. Here, we describe the complete sequential assignment and the tertiary structure in aqueous solution and propose preliminary quaternary structures relevant to CCL20 function. In addition, the CCL20 structure is compared with those of other chemokines and with the β-defensin 2 (BD2)[1] structure (37). On the basis of

* This work was supported by Dirección General de Investigación Científica y Técnica Grant PB98-0677 (Spain). The Departamento de Inmunología y Oncología was founded and is supported by the Consejo Superior de Investigaciones Científicas and Pharmacia Corporation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 1ha6) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

¶ To whom correspondence should be addressed. Tel.: 34 91 561 94 00; Fax: 34 91 564 24 31; E-mail: marta@malika.iem.csic.es.

[1] The abbreviations used are: BD1 and BD2, β-defensin 1 and 2, respectively; HPLC, high pressure liquid chromatography; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; r.m.s., root mean square; RANTES, regulated on activation normal T cell expressed and secreted; ASA, accessible surface area.

these results, we have identified some of the CCL20 structural motifs that could mediate its specific binding to CCR6. Also, we described structural features in BD2 that could explain its binding to CCR6.

EXPERIMENTAL PROCEDURES

Chemical Synthesis of CCL20—Mouse CCL20 (8) (ASNYDCCLSY IQTPLPSRAI VGFTRQMADE ACDINAIIFH TKKRKSVCAD PKQN-WVKRAV NLLSLRVKKM) was synthesized on a fully automatic peptide synthesizer (Applied Biosystems 430) by Fmoc (*N*-(9-fluorenyl)methoxycarbonyl) chemistry, according to standard protocols (38). After synthesis, protecting side chain groups were removed, and the peptide was cleaved from the resin following the method of King *et al.* (39). Folding and disulfide bridge formation of the HPLC product was performed as described elsewhere (40). The purity and composition of the folded polypeptide was confirmed by HPLC on a C-18 Nucleosil 120 analytical column (Tracer), by amino acid analysis using a Beckman 6300 amino acid analyzer and by mass spectrometry using a matrix-assisted laser desorption and ionization time-of-flight mass spectrometer, model Reflex III (Bruker-Franzen Analytic GmbH, Bremen, Germany) equipped with the SCOUT™ source in positive ion reflector mode using delayed extraction and AutoXecute™ acquisition software. Ion acceleration voltage was 25 kV. The equipment was externally calibrated employing protonated mass signals from a peptide mixture covering the 1000–12,300 *m/z* range. The experimental molecular mass obtained was 7952.6 daltons (calculated molecular mass, 7948 Da).

Sedimentation Equilibrium—Sedimentation equilibrium experiments were performed in a Beckman Optima XL-A ultracentrifuge using a Ti60 rotor and double sector centerpieces of Epon-Charcoal (optical pathlength 12 or 4 mm, depending upon peptide concentration). Samples of CCL20 in the concentration range from 0.15 to 0.02 mM and equilibrated in water with 200 mM NaCl, at pH 3.5 and 5.5, were centrifuged at 33,000 rpm and 20 °C. Radial scans at different wavelengths (from 280 to 300 nm) were taken every 2 h until equilibrium was reached, as judged by there being no difference between two successive scans. The weight average molecular masses of CCL20 were determined by using the program EQASSOC (Beckman) (41) with the partial specific volume calculated from its amino acid composition (42).

NMR Samples—NMR experiments were performed employing a ~1 mM sample at 25, 35, and 45 °C, pH 3.5 and 5.5 (uncorrected for deuterium isotope effects), containing 200 mM NaCl in either H_2O/D_2O (90%/10%, v/v) or D_2O. Sodium trimethylsilyl [2,2,3,3-$^2H_4$]propionate was used as the internal chemical shift reference.

NMR Spectroscopy—All spectra were acquired in the phase-sensitive mode with time-proportional phase incrementation mode (43) on a Bruker AMX-600 spectrometer equipped with a triple resonance probe and pulsed field gradients. Pulse gradients were used to suppress the water signal and undesired coherence pathways. ^1H homonuclear COSY (44), TOCSY with a mixing time of 60 ms (45), and NOESY (46) spectra with 80- and 150-ms mixing times were recorded by standard methods with water suppression achieved by selective presaturation of the water signal or by including the WATERGATE module (47) in the original pulse sequences. The size of the acquisition data matrix was 2048 × 512 words in f_2 and f_1, respectively. Before Fourier transformation, the two-dimensional data matrix was multiplied by a phase-shifted sine bell or square sine bell window function in both dimensions. The corresponding shift was optimized in every experiment. Base-line correction was applied in both dimensions. All spectra were processed and analyzed using XWINNMR (Bruker, Analytic Messtechnik GmbH, Ettlingen, Germany) and ANSIG (48) software on an IRIS Indigo work station.

Restraint Assignment—^1H NMR resonances were assigned using standard sequential assignment procedures (49). Spin systems were identified by analysis and comparison of the COSY and TOCSY spectra. The through-space connectivities were then determined using the NOESY spectra. The use of three temperatures and two pH values to record the spectra (25, 35, and 45 °C at pH 3.5 and 25 and 35 °C at pH 5.5) allowed us to resolve ambiguities that arose from chemical shift degeneracies.

The connectivities from NOESY spectra recorded at pH 3.5 were used as input for the structure calculation. After integration, the peak volumes and chemical shift lists were output to a DYANA-compatible format (50). For each NOESY spectrum, the automatic calibration method implemented in DYANA was run to transform the peak volumes into distances. The upper limits obtained were used without modification for the initial structure calculations. The assignment of cross-peaks in the NOESY spectra was carried out in several steps. A preliminary NOE data set was used to calculate initial structures with DYANA. The resulting conformers with lowest energies were used to assign additional peaks in the NOESY spectra. The cycle of calculations and assignments was repeated until no further assignments were possible. The quality of the restrictions was checked by analyzing the restraint violations of the calculated conformers. The NOE cross-peaks corresponding to restrictions that were consistently violated in a significant number of structures were checked for possible overlap, and the corresponding restraints were consequently modified. This cycle was repeated until no consistent violation was detected.

All structure calculations included the disulfide bond restraints Cys^6–Cys^{32} and Cys^7–Cys^{48} previously determined for CCL20, and in each case, these were fixed to a distance of 2.02 ± 0.1 Å. Slowly exchanging amide protons were identified in a series of TOCSY spectra recorded after dissolving fully protonated protein in D_2O. Hydrogen bond restraints were included in the later rounds of the structure calculation when the amide proton signal persisted in the D_2O spectrum and if an unambiguous partner was present in at least one-third of the preliminary structures. Each hydrogen bond was specified by two distance restraints (HN–O distance of 1.7–2.3 Å and N–O distance of 2.4–3.3 Å). In the initial stages of the calculations, ϕ angles were restrained to the range 0–180°, except for Gly and Asn residues. These restrictions were removed in the last steps of the calculation.

Structure Calculations and Analysis—The three-dimensional structure of CCL20 was determined using the program DYANA (50). According to the standard procedure, default weighting factors were attributed to the different restraint categories for different types of protons. Pseudoatom corrections were added to those upper limit distances involving stereospecific atoms that were not stereospecifically assigned. All peptide bonds were fixed as *trans*. All histidine, arginine, and lysine residues were regarded as positively charged, while the glutamate and aspartate side chains were treated as negatively charged. We have used a simulated annealing protocol in which 60 randomly generated structures were cooled down during 5000 torsion angle dynamics cycles to a final temperature of 0 K. Then a second simulated annealing step of 800 torsion angle dynamics cycles was done at low temperature with the minimized structures.

Following the torsion angle dynamics calculation, the 20 conformers with the lowest target function values were subjected to restrained energy minimization using the GROMOS force field (51). The resulting 20 energy minimized conformers represent the solution structure of CCL20.

The program MOLMOL (52) was used to visualize and globally characterize the structures. All r.m.s. deviation values reported here are mean pairwise r.m.s. deviation. Global r.m.s. deviation values are obtained from the best superposition of all considered residues.

RESULTS

CCL20 Is in an Association Equilibrium—Previously, several chemokines have been found to form dimers or higher order oligomers (53). To determine whether CCL20 oligomerizes in conditions similar to those used for NMR measurements, equilibrium sedimentation studies were performed. The determined mass of CCL20 in 200 mM NaCl at 20 °C, pH 5.5, and a variety of concentrations, shows a wide molecular mass distribution between 8.2 and 11.7 kDa (monomers' calculated molecular mass, 7948 Da). Similar results were found in the solution studies of CCL11/eotaxin (32), which showed that low pH values favor the monomeric species in the equilibrium. To test if this was also the case for CCL20, its molecular mass was also determined for samples at pH 3.5, and the result was similar to that mentioned above for pH 5.5. The equilibrium was also visible in the NMR spectra under several different conditions. At pH 5.5 and 25–35 °C, several chemical exchange peaks were clearly visible on TOCSY and NOESY spectra, confirming the existence of slow exchange (on the NMR time scale) of nuclei between different environments. The kinetics of the equilibrium is, however, pH-dependent, and is more rapid at lower pH values. Consequently, at pH 3.5, only a single set of signals, corresponding to the average of the populations present in solution, is observed in the NMR spectra.

Assignments—Resonance assignments were achieved at pH 3.5 and pH 5.5 using standard two-dimensional sequential

assignment techniques. At pH 5.5, the assignment process was hindered by the presence of two sets of signals and significant line broadening of some specific resonances. Consequently, signals corresponding to the amide protons of Phe^{23}, Arg^{25}, Gln^{26}, Met^{27}, Arg^{44}, Lys^{45}, Lys^{57}, Leu^{65}, Arg^{66}, Lys^{69}, and Met^{70} were not observable at this pH. In addition, exchange cross-peaks were identified for some resonances (*e.g.* Phe^{23} H_α (5.84 and 5.50 ppm) and Thr^{41} H_α (5.71 and 5.58 ppm)). These data indicate that these residues lie in two different magnetic environments and suggest, as observed for other chemokines (32), that these regions may be involved in forming quaternary contacts. At pH 3.5, all spin systems could be identified, and sequential cross-correlations were established on the basis of H_α-HN (H_α-H_δ for Pro), H_β-HN, and/or HN-HN NOEs observed between adjacent residues of the protein. For all residues, at least one sequential NOE could be unambiguously assigned with the exception of Leu^{62}-Leu^{63} due to signal overlap. The ^1H assignments, with the exception of rapidly exchanging protons, are almost complete. These assignments include all of the protons of the backbone and side chains excluding one H_γ in Pro^{51}, due to overlap, and the Met^{70} methyl group. Met^{70} is the C-terminal residue, and the high flexibility typical of such residues probably explains why its methyl group is not observed. The ^1H chemical shifts of murine CCL20 chemokine (pH 3.5, 35 °C) have been deposited in the BioMagResBank under accession number 2465.

The intensities of the main chain NOEs, the observed short and long range NOEs, and the conformational chemical shift of the $^1H_\alpha$ resonances indicate that the secondary structure of mouse CCL20 in solution consists of a single turn of helix, three β-strands, and a C-terminal α-helix, as has been observed in previous chemokine structures (17, 18). A total of 20 amide protons show protection after 15 min of deuterium exchange. The location of the protected amide protons shows broad agreement with the limits of secondary structure determined by the NOE data. This is similar to the hydrogen exchange pattern of other chemokines (54). The chemical shift index (55) and a schematic representation of the antiparallel β-sheet as determined from the NOE connectivities are shown in Fig. 1.

Calculation of the Solution Structure of Monomeric CCL20—The detailed analysis of the NOESY spectra of CCL20 at pH 3.5, at different protein concentrations and temperatures, did not reveal NOEs that could be attributed to intermonomer contacts, not even from residues at the N terminus or in the first β-strand. Accordingly, all of the assigned NOESY cross-peaks have been used in this work to build a monomeric model for CCL20 compatible with the experimental data. The input restraints used for the final round of structure calculations are listed in Table I. A total 1349 NOEs were collected and translated into 1004 relevant distance restraints, and 40 hydrogen bond restraints (two for each hydrogen bond; see "Experimental Procedures") were derived from the slow proton/deuterium exchange rates. The average number of restrictions per residue was 14.9 considering all residues and 17.1 for the well ordered residues (positions 6–66). As expected for a globular protein, the restrictions are well distributed throughout the structure (Fig. 2). The NMR structure of CCL20 is represented by 20 conformers with an average pairwise r.m.s. deviation of 1.70 and 2.31 Å for the backbone and all heavy atoms, respectively (Table II). When the N- and C-terminal ends (positions 1–5 and 67–70, respectively) are excluded from the comparison, the average pairwise r.m.s. deviation drops to 0.53 Å for the backbone and 1.36 Å for all heavy atoms. As is generally observed, the elements of secondary structure are the best defined regions of the protein. The loops connecting the different elements are also relatively well defined, and only the longest loop

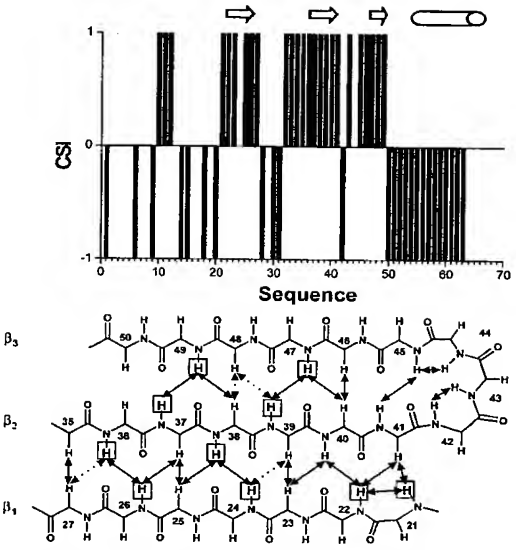

FIG. 1. *Top*, chemical shift index (*CSI*) (55) of CCL20 based on the chemical shifts at 35 °C and pH 3.5. Regions in β-strand or α-helix are depicted by *arrows* and a *cylinder*, respectively. *Bottom*, schematic representation of the antiparallel β-sheet of CCL20 as determined from the NOE connectivities. Observed interstrand and sequential HN-HN NOEs are indicated by *solid arrows*. Slow exchanging amide protons are *boxed*.

TABLE I
Summary of input restraints

Restraints	Number
Distance restraints	1004
Intraresidue ($i - j = 0$)	200
Sequential ($\|i - j\| = 1$)	239
Short range ($\|i - j\| < 5$)	171
Long range ($\|i - j\| > 5$)	394
Hydrogen bonds	20 × 2
Total restraints	1044
Average restraints per residue (1–70)	14.9
Average restraints per residue (6–66)	17.1

(positions 27–34) exhibits residues with r.m.s. deviation values slightly greater than 1.0 Å (Fig. 2). When this region is excluded from the comparison, r.m.s. deviation values decrease but are not substantially lower than those reported for residues 6–66. The quality of the final family of 20 structures is good, with no average NOE violations greater than 0.4 Å. Most of the backbone torsion angles fall in the most favored regions of Ramachandran space. The backbone conformation of the 20 structures are shown superimposed in Fig. 3. Atomic coordinates for the lowest energy 20 converged structures have been deposited in the Protein Data Bank with accession code 1ha6.

Description of the Structure—The structure of CCL20 has the general features previously determined for other chemokines. The N-terminal region (residues 1–4) preceding the CC motif is largely disordered with r.m.s. deviation values of 8.4–1.8 Å (Fig. 2). However, there are long distance NOEs between side chain protons of Asn^3 and Tyr^4 and those of Asp^{33} and Ile^{34}, indicating that these N-terminal residues are in close proximity to the protein core a significant fraction of the time. The N-loop region is bent in the CC motif (residues 6–7) and then extends to a turn of 3_{10} helix (residues 17–19). This 3_{10} helix, which is well defined in all of the structures, contains hydrogen bonds linking the Ala^{19} and Ile^{20} amide protons to the carbonyl groups of Pro^{16} and Ser^{17}, respectively. This turn motif is also present in other chemokine structures (36). The helical turn is followed by the first β-strand (residues 22–26) of a three-

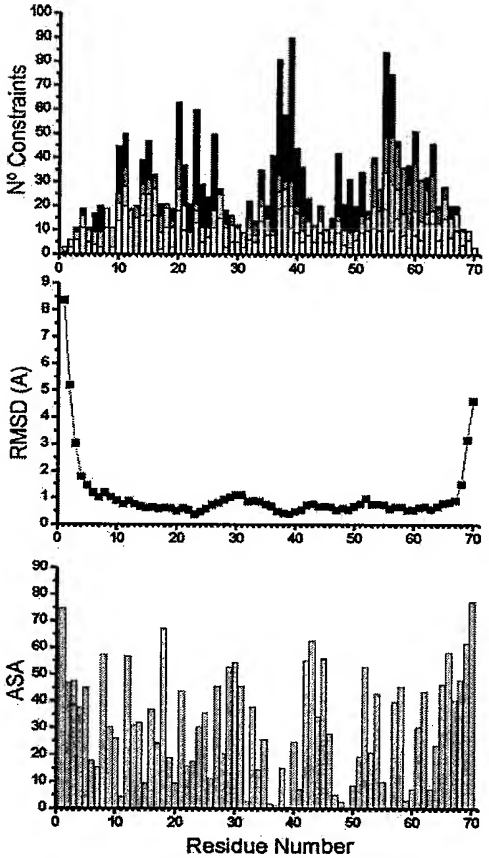

FIG. 2. *Top*, number of NOE constraints per residue used in the CCL20 structure calculation. NOEs between residues i and j are categorized as intraresidue ($i = j$, *black bar*), sequential ($|i - j| = 1$, *dark gray bar*), medium range ($1 < |i - j| = 4$, *light gray bar*), and long range ($|i - j| < 4$, *white bar*). *Middle*, r.m.s. pairwise deviation (Å) of backbone atoms between 20 structures of CCL20. *Bottom*, mean accessible surface area (%) per residue over the 20 solution structures of CCL20.

TABLE II
Summary of NMR structure calculations

Residual constraint violations	Sum	Maximum
	Å	
Upper limits	9.5	0.4
Energy terms	**Average**	**Range**
	kJ · mol⁻¹	
Total energy	−1990	−1408 to −2414
Lennard-Jones energy	−1347	−860 to −1678
NOE term	32	20–40
Pairwise r.m.s. deviation	**Backbone**	**All heavy atoms**
	Å	
All residues (1–70)	1.70 ± 0.56	2.31 ± 0.42
Residues 6–66	0.53 ± 0.12	1.36 ± 0.13
Residues 6–26, 35–66	0.48 ± 0.10	1.31 ± 0.17

stranded antiparallel β-sheet. Residues 36–40 and 47–49 comprise the second and third β-strand, respectively. As mentioned before, the loop connecting strands 1 and 2 is not as well defined as the rest of the molecule, whereas strands 2 and 3 are linked by ordered residues. Finally, strand 3 is connected by a turn to a regular α-helix extending from residue 54 to 66. The C-terminal residues show no long range NOE with the rest of the protein and are highly disordered, with large values of r.m.s. deviation, 1.5–4.6 Å.

FIG. 3. **Stereoscopic views of the CCL20 solution structure.** Superposition of the 20 structures with the N terminus on the *right* and the C terminus on the *left side* and α-helix face (*top*) and N-loop face (*middle*). *Bottom*, *ribbon* representation of the lowest energy CCL20 structure in the N-loop face orientation.

The α-helix is packed roughly orthogonal against the β-sheet axis. A significant number of hydrophobic interactions are formed between the sheet (Phe^{23}, Phe^{39}, and Val^{47}) and one side of the amphipathic α-helix (Trp^{55}, Val^{56}, Val^{60}, and Leu^{63}). In addition, residues Ile^{11}, Pro^{14}, and Ile^{15} in the N-terminal loop and Ile^{20} following the 3_{10} helical turn also make nonpolar contacts with each other and with Phe^{39}, Pro^{51}, and Trp^{55} in the core. The residues forming this pattern of hydrophobic contacts in CCL20 are well conserved throughout the entire chemokine family and have been postulated to be responsible for the structural similarities between the monomeric folds of CXC and CC chemokine subfamilies (56). Fig. 4 shows some details of the structure, the amphipathic α-helix, and part of the hydrophobic core. The α-helix is depicted with the hydrophobic face oriented toward the sheet, on the left, and the solvent-exposed hydrophilic face on the right.

The N-terminal region of the structure consists of a long loop that packs against the face of the β-sheet not covered by the α-helix. Disulfide bonds and two intramolecular hydrogen bonds, HN Ile^{11}–O Cys^{48} and HN Asp^{50}–OH Tyr^{10}, not specifically included in the calculations, link this loop to the β-sheet. The interface between the N-terminal region and the β-sheet appears to be further stabilized by hydrophobic contacts between the sidechains of Tyr^{4}, Tyr^{10}, Gln^{26}, Ile^{34}, and Ala^{49}. These contacts are mainly responsible for the conformation and the orientation of the N terminus of the protein, which protrudes from the structure in the opposite direction from the C terminus. Finally, the 3_{10} turn is attached to the β-sheet by means of a hydrogen bond, which was not explicitly included in the calculation, between the carbonyl oxygen of Ala^{19} and the amide proton of Lys^{42}.

An analysis of the accessible surface (Fig. 2) shows that, in general, residues exhibiting low values of accessible surface area (ASA) are located in structural elements, participate in

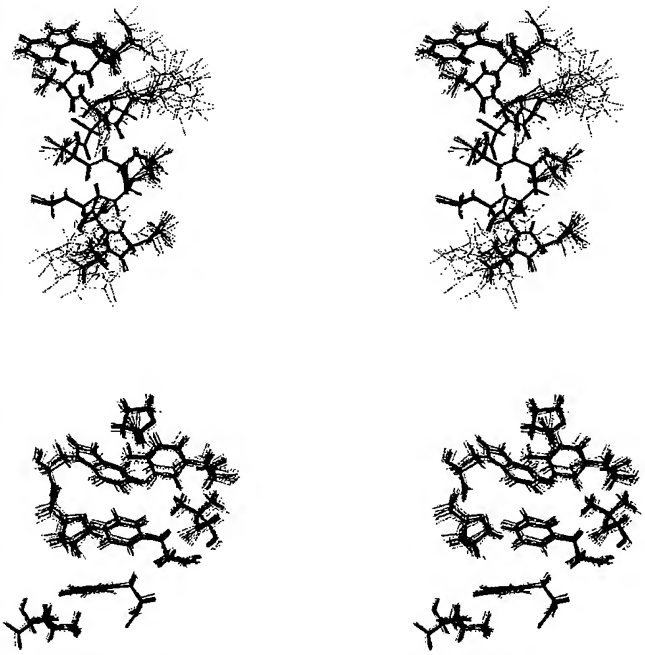

FIG. 4. **Stereoscopic view of the amphipatic C-terminal α-helix and the hydrophobic core of the solution structure of CCL20.** *Top*, the α-helix is represented by the best superposition of the backbone atoms of the Asn[54]–Arg[66] sequence. The *left side* of the helix is hydrophobic, and the *right side* of the helix is hydrophilic. *Bottom*, part of the hydrophobic core composed by side chains of residues belonging to the α-helix (Trp[55], Val[60], Leu[63]), the β-sheet (Phe[23], Phe[39], Val[47]), the N-loop (Pro[14], Tyr[10]), and Pro[51].

the internal hydrophobic core, and show a well defined conformation for the entire side chain. In contrast, charged and polar residues show larger values of ASA and are less well defined. Exceptions to these general rules are the hydrophobic Tyr[4], Leu[8], Val[21], Leu[62], and Leu[65], which are quite exposed (ASA, 40–60%), and well ordered.

DISCUSSION

Comparison with CCL7 Global Fold—Several CC and CXC chemokine structures have been solved by NMR and x-ray crystallography in recent years (17–36). The available data clearly indicate that the monomeric structure of all chemokines is very similar but that they show a variety of association states. Monomers share a common topology; all contain a 3_{10} helix followed by a three-stranded antiparallel β-sheet and a C-terminal α-helix. In particular, structural comparison of CCL20 with other CC chemokines produces superposition of the secondary elements with low r.m.s. deviation values indicative of structural similarity: CCL5/RANTES (26), 1.8 Å; CCL4/MIP-1β (22), 1.7 Å; CCL7/MCP-3 (31), 1.6 Å; and CCL2/MCP-1 (27), CCL11/eotaxin (32), and CCL1/I-309 (36), 1.4 Å. Despite this similarity, there are a few structural differences between structures that may be relevant for activity and provide some clues on the basis of receptor specificity.

Among chemokines with a CC motif whose three-dimensional structures have been solved, CCL20 structure is most similar to CCL7 (31), with which it shares the same N-terminal orientation. The cores of the CCL20 and CCL7 monomers superimpose relatively well with the exception of the 30's loop, where remarkable differences are observed. The closed end of this loop points toward the helix face in the case of CCL7, whereas in CCL20 it lies near the N-terminal loop. This different orientation is due to different packing interactions with the hydrophobic residues following the CC motif (*i.e.* Tyr[10] in CCL20 and Phe[15] in CCL7). The importance of the orientation of this aromatic position, which appears to play a role in concerted motions of the entire 30's loop and the N terminus (via the disulfide bond), has been discussed previously (32). Both aromatic residues are relatively solvent-exposed, and the phenolic OH of Tyr[10] in CCL20 makes a hydrogen bond with the amide proton of Asp[50] located at the base of the helix. Phe[15] in CCL7 is obviously unable to form such an interaction. This hydrogen bond results in a distinct side chain orientation which forms very different hydrophobic interactions. Consequently, Phe[15](CCL7) and Tyr[10](CCL20) cannot be considered to be structurally equivalent units. In addition to this Tyr/Phe substitution, there is another structural difference between CCL7 and CCL20: the relative position of the C terminus with respect to the sheet. This different orientation is due to the second Tyr/Phe swap. In CCL7, the aromatic residue in the first strand is Tyr[28], whereas in CCL20, it is Phe[23]. Tyr[28] links the sheet to the C-terminal region of the helix by a hydrogen bond between its OH and the backbone O of Lys 70, bending the structure at the C terminus. The lack of this interaction in CCL20 produces a slightly different orientation at the C terminus. Some other differences between these structures possibly related with activity and receptor binding of CCL20 are discussed below.

Implications for Dimerization—Ultracentrifugation analysis shows that CCL20 exists in an association, presumably monomer-dimer, equilibrium. The limitations of homonuclear NMR methods make it very difficult to detect intermolecular contacts that could account for a multimeric structure. The possible intersubunit NOEs may be under strong peaks, below the noise level or not detected because of the low population of these states. However, additional structural NMR data other than NOE, namely the presence of chemical exchange peaks and the signal loss because of broadening, clearly prove the presence of supramolecular structures. Therefore, if this associated state of CCL20 present under our experimental conditions were a dimer, we would propose that it is formed by association of the first β-strand (β_1) of each monomer to form a six-stranded antiparallel sheet. This structure is consistent with the monomer geometry and all of the NMR evidence in hand and has been observed in both CXC and CC chemokines, like CCL7 and CCL2 (30, 31, 57). It has been shown that only a few conserved residues, characteristic of each chemokine subfamily, are involved in dimer stabilization (56). In the proposed dimeric structure of CCL20, Phe[23] (strand 1), Arg[25] (strand 1), Ser[64] (helix), and Leu[63] (helix) could account for the quaternary contacts. In a similar way, the homologous residues Tyr[28], Arg[30], Asp[68], and Leu[67] form the dimer contacts in CCL7 (31).

Structure-Function of CCL20—Several studies have characterized the interaction between chemokines and cellular receptors (1, 4, 6). Chemokines bind to negatively charged sequences at the N termini of the cellular chemokine receptors and also to charged residues on some of the external loops. It is also known that amino acid residues preceding the first cysteine and the N-loop are important for binding and activating the receptor (58). Although chemokines share significant sequence similarity in regions of common secondary structure, the N terminus of each chemokine is largely distinct, even within members of the same subfamily (59). However, it is common for chemokine sequences to contain an acidic residue among the first residues. For example, studies done on the ELR motif near the N terminus of CXC chemokines (60) suggest that this motif is involved in hydrophobic and electrostatic contacts with extracellular loops 3 and 4 of the receptor. It is well known that most chemokines can bind to more than one receptor in a distinct

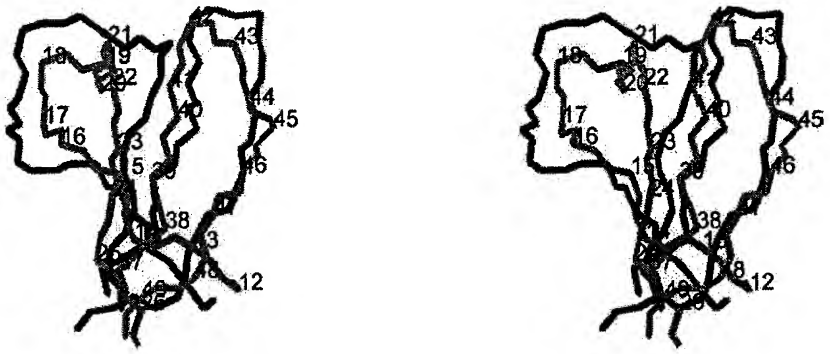

FIG. 5. **Stereoscopic view of the groove between the N-loop and the β_2-β_3 hairpin.** Backbone trace of the superposition of CCL20 (residues 12–26 and 37–49; *red*) and CCL7 (31) (residues 16–30 and 41–53; *blue*). α-Carbons are labeled in the CCL20 trace.

and specific manner and that one residue at the N terminus (*e.g.* Asp[6] in CCL5) can be involved in the activation of more than one receptor (61). To make these interactions, such key residues need to be located in flexible regions that can be adjusted to fit the distinct surface topology of different receptors.

The existence of an Asp-Cys-Cys-Leu (DCCL) motif in the N-end of CCL20, shared only by human and murine CCL19/ELC, CCL21/SLC, and CCL25/TECK and murine CCL9–10/C10-like has been reported (8), although the contributions of this motif to function are unknown. Recently, human CCL20 variants that have short deletions in the flexible N termini but maintain the DCCL motif and/or the C termini (positions 2–70, 1–69, 1–67, 1–64, 2–69, 2–67, and 2–64) were all shown to be active in intracytoplasmic calcium mobilization as well as chemotactic assays (62). Inspection of the CCL20 structure shows that residues Asp[5] and Leu[8] have a very well defined conformation, with both side chains oriented in the same direction and largely exposed to the solvent. These characteristics suggest that these residues are good candidates to make specific interactions with the receptor. Asp[5] and Leu[8] flank the CC motif held against the sheet by two S–S bonds. Consequently, this region should be quite rigid in contrast to the ELR motif present in some CXC chemokines, which is located far from disulfide bonds. This difference probably has implications for receptor specificity. We propose that the rigidity of the DCCL motif limits its ability to adopt alternate conformations that would allow it to recognize and bind different receptors. Human chemokines with a DCCL motif bind to two receptors (one specific (CCR7 for CCL19 and CCL21 or CCR9 for CCL25) and the another one shared (CCR11)). Therefore, they also bind selectively, although not so much as CCL20.

However, the presence of the DCCL motif in these other non-CCR6-binding chemokines strongly suggests that other parts of CCL20 play a role in its specific binding to CCR6. In other chemokines, such as RANTES (63), multiple structural regions are required for specificity. Structures of both CC (54) and CXC (59) chemokines complexed with peptide models of their receptors show that the peptide receptor mimics bind in a groove formed by the N-loop and the β_2-β_3 hairpin of the chemokines. Recently, a model for chemokine/receptor recognition has been proposed based on the binding of residue motifs and correct binding to the chemokine groove (59).

While many of the residues defining the edges of this groove channel are charged or polar, the floor of the groove is formed predominantly by hydrophobic side chains. Interestingly, remarkable differences are observed in the size of the groove; in CCL20, it is much more narrow than in CCL7 or CXCL8. Fig. 5 shows a superposition of CCL20 and CCL7 focusing on this region. Both structures match the orientation of the β_2-β_3 hairpin, but significant discrepancies are observed in the conformation of the N-loops that narrow the groove in CCL20.

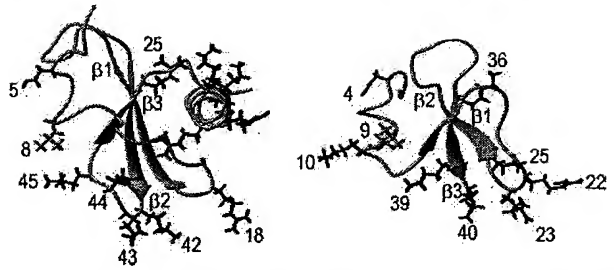

FIG. 6. ***Ribbon* representation of the chemokine CCL20 (*left*) and the β-defensin BD2 (37) (*right*).** β-Sheets are represented by *arrows*; arginine and lysine residues are shown in *blue*; Asp[6] in CCL20 and Asp[4] in BD2 are depicted in *red*; and Leu[8] in CCL20 and Leu[9] in BD2 are shown in *green*.

Although it is difficult to precisely quantify the width of the groove, an idea of its dimensions can be obtained by comparing the corresponding backbone positions in CCL20, CCL7, and CXCL8. Thus, whereas the distance between α-carbons of Lys[42] (in the middle of the β_2-β_3 hairpin) and Ile[20] (on the top of the N-loop in Fig. 5) is about 6 Å in CCL20, it is 11 Å between equivalent positions in CCL7 (Leu[47]–Arg[25]) and 9 Å in CXCL8 (Asp[45]–Lys[23]). Other related distances between both segments are always smaller in CCL20 than in the other proteins (*e.g.* Arg[44]–Thr[13], 11 Å in CCL20; Lys[46]–Phe[15], 18 Å in CCL7; and Glu[48]–Phe[17], 14 Å in CXCL8). It is tempting to speculate that the narrow channel in CCL20 can only fit a very specific counterpart and contributes to its highly specific receptor binding. We therefore propose that the narrow groove and the rigidity at the DCCL motif can account for the singularly specific binding between CCL20 and its CCR6 receptor.

Comparison with Human β-Defensins—Recently, human β-defensins have been shown to bind to the CCR6 receptor with significant affinity (16). Although β-defensins show no sequence homology with chemokines, it has been postulated that they could have similar tertiary CCR6 binding sites and act as "microchemokines." To date, two β-defensins, BD1 and BD2, have been described in humans, and the structure of the latter has been solved by crystallography (37). The monomeric structure of BD2, stabilized by three disulfide bridges, is composed of a three-stranded β-sheet (residues 14–16, 25–28, and 36–39) and a short α-helical region (residues 5–10) held against the sheet by a S–S bond. A comparison between this β-defensin and the CCL20 backbone structure is depicted in Fig. 6, showing their positively charged side chains, the motif DCCL in CCL20, and Asp[4] and Leu[9] in BD2. As can easily be seen, β-defensin can structurally be considered as a simplified form of CCL20, with a truncated N-terminal end and lacking the C-terminal helix. In the orientation presented in Fig. 6, it is evident that the position of the β-sheets is similar in both molecules, despite their different topology (β_1 up, β_2 down, β_3 up, in CCL20, and

β_1 up, β_2 up, β_3 down in BD2). Moreover, the positively charged side chains (Arg^{18}, Arg^{25}, Lys^{42}, Lys^{43}, Arg^{44}, and Lys^{45} in CCL20 and Lys^{10}, Arg^{22}, Arg^{23}, Lys^{25}, Lys^{36}, Lys^{39}, and Lys^{40} in BD2) are distributed in a similar way. Thus, the β-sheet and the specific distribution of charged residues seem to be common features in both structures. We have already discussed the possible importance of the groove's width between the N-loop and the β_2-β_3 hairpin, and the DCCL motif for CCL20 specificity. Interestingly, BD2 has two similar residues, Asp^4 and Leu^9, placed in similar position and orientation as those of Asp^5 and Leu^8 in CCL20. If the DCCL motif plays a role in the binding of CCL20 to CCR6, the pair Asp^4–Leu^9 in BD2 could provide similar interactions. These residues are located in a regular α-helix structure strongly held against the protein core by a disulfide bridge. In addition, these residues are moderately conserved among β-defensin sequences (37); this supports the idea that specific interactions with CCR6 need appropriate groups in a fixed rigid orientation. Whereas a structural homology of the previously mentioned CCL20 groove is not obvious at first glance in BD2, a closer inspection reveals a region that could bind CCR6 in a similar fashion. The CCL20 groove is oriented face-on in Fig. 6; therefore, if BD2 binds CCR6 in a similar way as CCL20 does, by structural analogy the interacting region in BD2 would comprise β_1 (residues 14–16), the loop connecting β_1 and β_2 (residues 17–24), and residues 36–40. Strikingly, sequence alignment of the fragment 14–27 of BD2 (IHPVFCPRRYKGI) reveals a significant degree of similarity with the N-loop region 11–23 of CCL20 (IQTPLPSRAIVGF), and the charged end of the sequence of BD2 (36–40, KCCKK) corresponds to the charged sequence 42–45 (KKRK) in the β_2-β_3 hairpin of CCL20. These correspondences and the close structural similarity may provide BD2 with the necessary groups to interact specifically with the CCR6 receptor, although with less affinity than CCL20, as has been described (19).

Clusters of positive charges have been claimed to play a role in the activities of both defensins and chemokines (37, 60). We have compared the positive charge distribution of BD2 and CCL20 with other chemokines, looking for differences. We found that BD2 and CCL20 have a highly asymmetrical distribution of the positive charge (Fig. 6) that is also shared by other chemokines examined (not shown). This suggests that these charged clusters may play a common role in the activity of all of these proteins but are not important for determining their receptor binding specificity.

In summary, we have determined the structure of CCL20, the chemokine with the most selective known receptor binding properties. Our data suggest that this strict receptor selectivity is related to the rigid conformation of the DCCL motif as well as the narrow groove formed by the N-loop and the β_2-β_3 hairpin, compared with other chemokines. Some of these structural features are also found in the defensin BD2, thus supporting their importance for the binding of both CCL20 and BD2 to CCR6. The determination of the three-dimensional structure of CCL20 will aid the rational design of variants to modulate or interfere with its important biological functions.

Acknowledgments—We thank Jesús Vázquez and Fernando Barahona (Protein Chemistry Laboratory, Centro de Biología Molecular Severo Ochoa) for help in peptide synthesis.

REFERENCES

1. Rollins, B. J. (1997) *Blood* **90,** 909–928
2. Baggiolini, M. (1998) *Nature* **392,** 565–568
3. Luster, A. D. (1998) *N. Engl. J. Med.* **338,** 436–445
4. Zlotnik, A., and Yoshie, O. (2000) *Immunity* **12,** 121–127
5. Müller, A., Homey, B., Soto, H., Ge, N., Catron, D., Buchanan, M. E., McClanahan, T., Murphy, E., Yuan, W., Wagner, S. N., Barrera, J. L., Mohar, A., Verástegui, E., and Zlotnik, A. (2001) *Nature* **410,** 50–56
6. Murphy, P. M., Baggiolini, M., Charo, I. F., Hebert, C. A., Horuk, R., Matsushima, K., Miller, L. H., Oppenheim, J. J., and Power, C. A. (2000) *Pharmacol. Rev.* **52,** 145–176
7. Baba, M., Imai, T., Nishimura, M., Kakizaki, M., Takagi, S., Hieshima, K., Nomiyama, H., and Yoshie, O. (1997) *J. Biol. Chem.* **272,** 14893–14898
8. Varona, R., Zaballos, A., Gutiérrez, J., Martín, P., Roncal, F., Albar, J. P., Ardavín, C., and Márquez, G. (1998) *FEBS Lett.* **440,** 188–194
9. Tanaka, Y., Imai, T., Baba, M. Ishikawa, I., Uehira, M., Nomiyama, H., and Yoshie, O. (1999) *Eur. J. Immunol.* **29,** 633–642
10. Iwasaki, A., and Kelsall, B. (2000) *J. Exp. Med.* **191,** 1381–1394
11. Dieu, M. C., Vanbervliet, B., Vicari, A., Bridon, J. M., Oldham, E., Aït-Yahia, S., Briere, F., Zlotnik, A., Lebecque, S., and Caux, C. (1998) *J. Exp. Med.* **188,** 373–386
12. Carramolino, L., Kremer, L., Goya, I., Varona, R., Buesa, J. M., Gutiérrez, J., Zaballos, A., Martínez-A., C., and Márquez, G. (1999) *J. Leukocyte Biol.* **66,** 837–844
13. Liao, F., Rabin, R. L., Smith, C. S., Sharma, G., Nutman, T. B., and Farber, J. M. (1999) *J. Immunol.* **162,** 186–194
14. Cook, D. N., Prosser, D. M., Forster, R., Zhang, J., Kuklin, N. A., Abbondanzo, S. J., Niu, X. D., Chen, S. C., Manfra, D. J., Wiekowski, M. T., Sullivan, L. M., Smith, S. R., Greenberg, H. B., Narula, S. K., Lipp, M., and Lira, S. A. (2000) *Immunity* **12,** 495–503
15. Varona, R., Villares, R., Carramolino, L., Goya, I., Zaballos, A., Gutiérrez, J., Torres, M., Martínez-A., C., and Márquez, G. (2001) *J. Clin. Invest.* **107,** R37–R45
16. Yang, D., Chertov, O., Bykovskaia, S. N., Chen, Q., Buffo, M. J., Shogan, J., Anderson, M., Schöder, J. M., Wang, J. M., Howard, O. M. Z., and Oppenheim, J. J. (1999) *Science* **286,** 525–528
17. St. Charles, R., Walz, D. A., and Edwards, B. F. P. (1989) *J. Biol. Chem.* **264,** 2092–2099
18. Clore, G. M., Appella, E., Yamada, M., Matsushima, K., and Gronenborn, A. M. (1990) *Biochemistry* **29,** 1689–1696
19. Fairbrother, W. J., Reilly, D., Colby, T. J., Hesselgesser, J., and Horuk, R. (1994) *J. Mol. Biol.* **242,** 252–270
20. Hanzawa, H., Haruyama, H., Watanabe, K., and Tsurufuji, S. (1994) *FEBS Lett.* **354,** 207–212
21. Kim, K. S., Clark-Lewis, I., and Sykes, B. D. (1994) *J. Biol. Chem.* **269,** 32909–32915
22. Lodi, P. J., Garret, D. S., Kuszewski, J., Tsang, M. L. S., Weatherbee, J. A., Leonard, W. J., and Gronenborn, A. M. (1994) *Science* **263,** 1762–1767
23. Chung, C., Cooke, R. M., Proudfoot, A. E. I., and Wells, T. N. C. (1995) *Biochemistry* **34,** 5329–5342
24. Malkowski, M. G., Wu, J. Y., Lazar, J. B., Johnson, P. H., and Edwards, B. F. (1995) *J. Biol. Chem.* **270,** 7077–7087
25. Rajarathnam, K., Clark-Lewis, I., and Sykes, B. D. (1995) *Biochemistry* **34,** 12983–12990
26. Skelton, N. J., Aspiras, F., Ogez, J., and Schall, T. J. (1995) *Biochemistry* **34,** 5329–5342
27. Handel, T. M., and Domaille, P. J. (1996) *Biochemistry* **35,** 6569–6584
28. Kim, K. S., Rajarathnam, K., Clark-Lewis, I., and Sykes, B. D. (1996) *FEBS Lett.* **395,** 277–282
29. Crump M. P., Gong, J., Loetscher, P., Rajarathnam, K., Amara, A., Arenzana-Seisdedos, F., Virelizier, J., Baggiolini, M., Sykes, B. D., and Clark-Lewis, I. (1997) *EMBO J.* **16,** 6996–7007
30. Lubkowski, J., Bujacz, G., Boque, L., Domaille, P. J., Handel, T. M., and Wlodawer, A. (1997) *Nat. Struct. Biol.* **4,** 64–69
31. Meunier, S., Bernassau, J. M., Guillemot, J. M., Ferrara, P., and Darbon, H. (1997) *Biochemistry* **36,** 4412–4422
32. Crump M. P., Rajarathnam, K., Kim, K., Clark-Lewis, I., and Sykes, B. D. (1998) *J. Biol. Chem.* **273,** 22471–22479
33. Dealwis, C., Fernández, E. J., Thompson, D. A., Simon, R. J., Siani, M. A., and Lolis, E. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 6941–6946
34. Shao, W., Jerva, L. F., West, J., Lolis, E., and Schweitzer, B. I. (1998) *Biochemistry* **37,** 8303–8313
35. Mizoue, S., Bazan, J. F., Johnson, E. C., and Handel, T. M. (1999) *Biochemistry* **38,** 1402–1414
36. Keizer, D. W., Crump, M. P., Lee, T. W., Slupsky, C. M., Clark-Lewis, I., and Sykes, B. D. (2000) *Biochemistry* **39,** 6053–6059
37. Hoover, D. M., Rajashankar, K. R., Blumenthal, R., Puri, A., Oppenheim, J. J., Chertov, O., and Lubkowski, J. J. (2000) *J. Biol. Chem.* **275,** 32911–32918
38. Fields, G. B., and Noble, R. L. (1990) *Int. J. Pept. Protein Res.* **35,** 161–124
39. King, D. S., Fields, C. G., and Fields, G. B. (1990) *Int. J. Pept. Protein Res.* **36,** 255–266
40. Clark-Lewis, I., Vo, L., Owen, P., and Anderson, J. (1997) *Methods Enzymol.* **287,** 233–250
41. Minton, A. P. (1994) in *Modern Analytical Ultracentrifugation* (Schuster, T. M., and Laue, T. M., eds) pp. 81–93, Birkhauser, Boston
42. Laue, T. M., Shak, B. D., Ridgeway, T. M., and Pelletier, S. L. (1992) in *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding, S. E., Rowe, A. J., and Horton, J. C., eds) pp. 90–125, Royal Society of Chemistry, Cambridge
43. Marion, D., and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.* **113,** 967–974
44. Aue, W. P., Bartholdi, E., and Ernst, R. R. (1976) *J. Chem. Phys.* **64,** 2229–2246
45. Bax, A., and Davis, D. G. (1985) *J. Magn. Reson.* **65,** 355–360
46. Kumar, A., Ernst, R. R., and Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun.* **95,** 1–6
47. Piotto, M., Saudek, V., and Sklenár, V. (1992) *J. Biomol. NMR* **2,** 661–665
48. Kraulis, P. J., Domaille, P. J., Campbell-Burk, S. L., van Aken, T., and Laue, E. D. (1994) *Biochemistry* **33,** 3515–3531
49. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, Inc., New York
50. Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997) *J. Mol. Biol.* **273,** 283–298

51. van Gunsteren, W. F., and Berendsen, H. J. C. (1987) *Groningen Molecular Simulation (GROMOS) Library Manual*, Biomos, Groningen, The Netherlands
52. Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graph.* **14,** 51–55
53. Stone, M. J., and Mayer, K. L. (1999) in *Chemokines in Allergic Disease* (Rothenberg, M. E., ed) pp. 67–94, Marcel Dekker, New York
54. Mayer, K. L., and Stone, M. J. (2000) *Biochemistry* **39,** 8382–8395
55. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992) *Biochemistry* **31,** 1647–1651
56. Covell, D. G., Smythers, G. W., Gronenborn, A. M., and Clore, G. M. (1994) *Protein Sci.* **3,** 2064–2072
57. Clore, G. M., and Gronenborn, A. M. (1991) *J. Mol. Biol.* **217,** 611–620
58. Clark-Lewis, I., Kim, K. S., Rajarathnam, K., Gong, J. H., Dewald, B., Moser, B., Baggiolini, M., and Sykes, B. D. (1995) *J. Leukocyte Biol.* **57,** 703–711
59. Schall, T. J. (1991) *Cytokine* **3,** 165–183
60. Skelton, N. J., Quan, C., Reilly, D., and Lowman, H. (1999) *Structure Fold. Des.* **7,** 157–168
61. Pakianathan, D. R., Kuta, E. G., Artis, D. R., Skelton, N. J., and Hébert, C. A. (1997) *Biochemistry* **36,** 9642–9648
62. Schutyser, E., Struyf, S., Menten, P., Lenaerts, J. P., Conings, R., Put, W., Wuyts, A., Proost, P., and Van Damme, J. (2000) *J. Immunol.* **165,** 4470–4477
63. Arenzana-Seisdedos, F., Virelizier, J. L., Rousset, D., Clark-Lewis, I., Loetscher, P., Moser, B., and Baggiolini, M. (1996) *Nature* **383,** 400